



**Intelligence Community and Department of Defense
Content Discovery & Retrieval Integrated Project Team**

**IC/DoD SOAP Interface Specification
for CDR Retrieve**

Version 2.0

Approval Date: 3-OCT-2012

Distribution Notice

This document has been approved for Public Release and is available for use without restriction

REVISION/HISTORY

Doc Revision	Revised By	Revision Date	Revisions
0.1		26 February 2010	Initial draft for subgroup review.
1.0-20100312	CDR IPT	12 March 2010	Minor updates.
1.0-20100319	CDR IPT	19 March 2010	Minor updates.
1.0-20100327	CDR IPT	27 March 2010	Minor updates.
1.0-20100331	CDR IPT	31 March 2010	Minor updates.
1.0-20100331	Farley	29 April 2010	Tech Edits
2.0	Wigglesworth	17 February 2012	Harmonization

Table of Contents

1	Introduction.....	6
1.1	Service Overview	6
1.2	Scope.....	6
1.3	Artifact Overview	6
1.4	Notational Convention	8
1.5	Conformance.....	8
1.6	Namespaces.....	8
1.7	Security	9
2	Retrieve Service Interface	10
2.1	Retrieve Function.....	10
2.1.1	Preconditions.....	10
2.1.2	Input	10
2.1.3	Output	11
2.1.4	Post-Conditions.....	12
2.1.5	Fault Conditions.....	13
	References.....	14
	Appendix A. Mapping to Specification Framework	16
A.1.	Retrieve Request	16
A.2.	Retrieve Response.....	16

LIST OF FIGURES

Figure 1: CDR Architecture Model	7
Figure 2: Example Retrieve Input.....	11
Figure 3: Example Retrieve Output	12
Figure 4: Example Retrieve Output (MTOM).....	12
Figure 5: Example Retrieve SOAP Fault.....	14

LIST OF TABLES

Table 1: Referenced XML Namespaces	9
Table 2: Header Elements for Retrieve Requests	10
Table 3: Required Header Elements for Retrieve Responses	11
Table 4: List of Faults	13
Table 5: Retrieve Request Mapping to Specification Framework.....	16
Table 6: Retrieve Response Mapping to Specification Framework	16

1 Introduction

1.1 Service Overview

The Retrieve Component, as defined by the “Intelligence Community/Department of Defense (IC/DoD) Content Discovery and Retrieval (CDR) Specification Framework” [CDR-SF], is the primary mechanism for content consumers to access one or more specific content resources from content collections. This component provides a common service interface and behavioral model for IC and DoD content collections, enabling consumers to retrieve and initiate delivery of content resources. Specifically, the Retrieve Component provides a means to retrieve the native content described in the Search Component query results.

This specification defines requirements and provides guidance for the realization of the CDR Retrieve Component as a web service using SOAP¹, hereafter termed a Retrieve Service in this document. The content of this specification describes the Retrieve Service’s behavior, interface and other aspects in detail, providing enough information for Retrieve Service providers and consumers to create and use CDR-conformant Retrieve Services.

The Retrieve service exposes a single Retrieve function. While the function is often used in concert with retrieving results of a search, it may be used in general to process any compliant retrieve instructions. As discussed in CDR Specification Framework, a Retrieve Service’s results are the content resource. In the context of Retrieve, the content resource generally refers to the entire underlying record.

1.2 Scope

The Retrieve Component as defined supports the retrieval of a specified resource from a content collection and returning that content to the requestor.

1.3 Artifact Overview

This specification is a part of the set of specifications that define the concrete, implementation-specific guidance for the services defined under the auspices of the Content Discovery & Retrieval (CDR) Integrated Project Team (IPT). The CDR Reference Architecture [CDR-RA] prescribes an abstract-to-concrete model for the development of architecture elements and guidance for content discovery and retrieval. Each layer or tier of the model is intended to provide key aspects of the overall guidance to achieve the goals and objectives for joint DoD/IC content discovery and retrieval. The following graphic, discussed in detail within the CDR Reference Architecture, illustrates this model.

¹ SOAP is a protocol used by web services in the exchange of structured information. [SOAP]

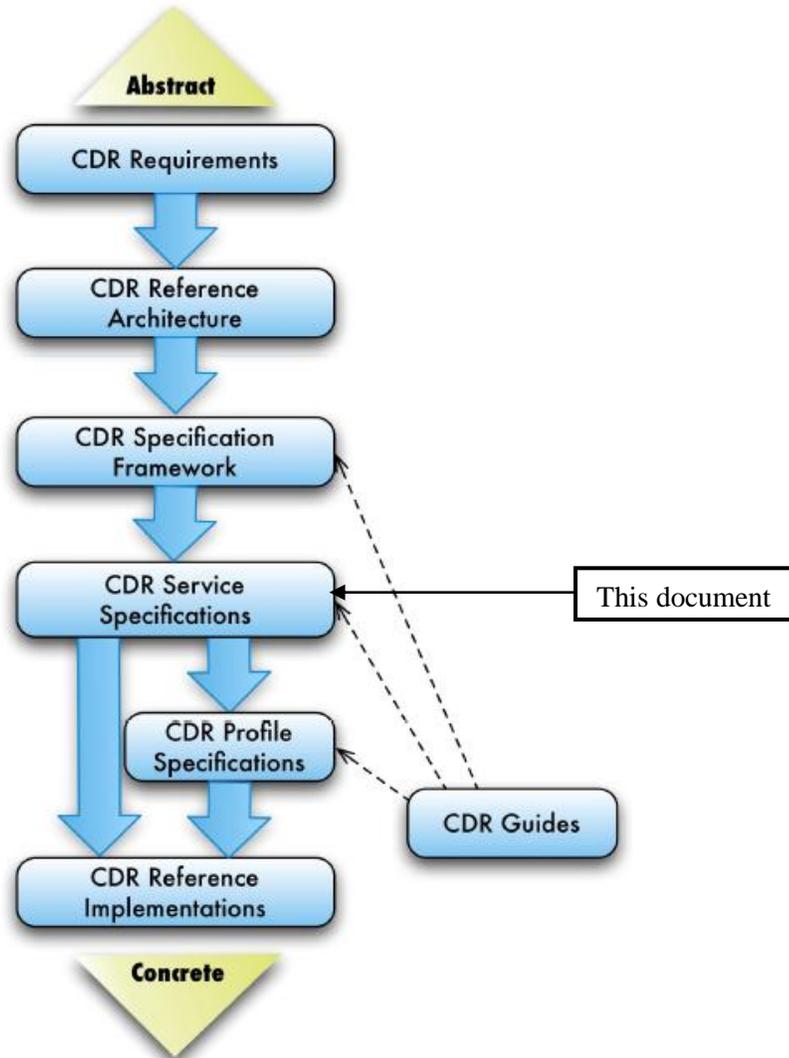


Figure 1: CDR Architecture Model

As illustrated in Figure 1, the CDR Specification Framework [CDR-SF] derives from the CDR Reference Architecture [CDR-RA] and describes behavior in terms of the capabilities, components, and usage patterns defined in the RA. Multiple CDR Service Specifications are derived from the CDR-SF, with separate specifications associated with the components of the architecture (e.g., Retrieve) and, for each service, separate specifications to address Representational State Transfer (REST) and SOAP implementations.

This specification provides guidance for implementing the CDR Retrieve Service as a SOAP Web Service. It is intended to parallel the corresponding REST specification, the IC/DoD Content Discovery & Retrieval REST Interface Specification for CDR Retrieve

[CDR-RR], as closely as possible, to minimize the difficulties in interoperating. Additional CDR Guides, Profile Specifications, or Reference Implementations may provide additional guidance on implementing this specification in a particular context.

1.4 Notational Convention

The key words "MUST," "MUST NOT," "REQUIRED," "SHALL," "SHALL NOT," "SHOULD," "SHOULD NOT," "RECOMMENDED," "MAY," and "OPTIONAL" in this specification are to be interpreted as described in the IETF RFC 2119. When these words are not capitalized, they are meant in their natural-language sense.

When describing concrete XML schemas and example XML documents, this specification uses XPath as the notational convention. Each member of an XML schema is described using an XPath notation (e.g., `/x:RootElement/x:ChildElement/@Attribute`). The use of `{any}` indicates the presence of an element wildcard (`<xs:any/>`). The use of `@{any}` indicates the presence of an attribute wildcard (`<xs:anyAttribute/>`).

Items contained in curly braces (`{item}`) are meant to indicate template or notional values to be replaced by actual values (without the use of curly braces) when in actual use.

Examples in this text are distinguished by a black border. These are meant to be illustrative and only one way that the described syntax can be used.

```
<atom:entry>  
  <atom:title>This is an example.</atom:title>  
</atom:entry>
```

Examples figured in this document are pseudo in that XML documents may not be valid complete XML documents and SOAP messages may not be valid complete SOAP messages, however, the examples is meant to demonstrate something specific in this document.

1.5 Conformance

This specification defines an interface to a Retrieve Service to which an implementation and a subsequent deployment **MUST** conform. A deployment is an instance of an implementation. For an implementation to conform to this Retrieve specification, it **MUST** adhere to all mandatory aspects of the specification.

1.6 Namespaces

Namespaces referenced in this document and the prefixes used to represent them are listed in the following table. The namespace prefix of any XML Qualified Name (QName) used in any example in this document should be interpreted using the information below.

Table 1: Referenced XML Namespaces

Prefix	URI	Description
soap	http://www.w3.org/2003/05/soap-envelope	SOAP 1.2 Envelope
cdrr	urn:cdr:retrieve:2.0	The CDR IPT Retrieve binding for SOAP implementations

1.7 Security

This specification does not directly address security concerns. It will be necessary for any implementation of this specification to address security concerns relevant to the systems with which they interact and the governance bodies. Several aspects of retrieve, to include returning only the content resource for which the requesting entity is authorized, should be addressed in the detailed security plan of an implementation, but are out of scope for this document.

2 Retrieve Service Interface

2.1 Retrieve Function

2.1.1 Preconditions

The following preconditions **MUST** be satisfied if the retrieve function is to correctly process input and generate results and post-conditions as described.

1. The requester is authenticated and authorized according to applicable policy requirements for this function.
2. The resource exists and can be retrieved.

2.1.2 Input

The input to the CDR Retrieve Service **MUST** be a valid SOAP² message that meets criteria identified in this section. The input should be directed to the SOAP Endpoint address identified by the implementer.

2.1.2.1 Header

The header of the SOAP message **MUST** contain the action element, as defined in WS-Addressing [WS-A]. The purpose of this element is to convey to the service which behavior to invoke. The header also contains the `cdrr:Identifier`. Additional elements, such as other WS-Addressing elements, **MAY** be added to the SOAP header.

Table 2: Header Elements for Retrieve Requests

Element Name Description	Support
/wsa:Action This element (whose content is of type <code>xs:anyURI</code>) conveys the value of the [action] property and indicates to a web service which operation should be invoked.	MUST be supported by Service. MUST be provided by consumer with a value of <code>urn:cdr:retrieve:2.0</code>
/cdrr:Identifier This element contains the unique identifier for the requested resource being retrieved.	MUST be supported by Service. MUST be provided by consumer
/cdrr:RetrieveProperties This element contains additional optional properties for the Retrieve function.	MAY be supported by Service MAY be provided by consumer

Additional extension attributes **MAY** be supported by the Retrieve Service and/or provided by the consumer to convey additional Retrieve properties. However, any additional attributes supported by the Retrieve Service **MUST** be **OPTIONAL** for the consumer to provide. If the consumer provides extension attributes that the Retrieve Service does not support, these **MUST** be ignored.

² Consult the relevant standards registry (such as the ICSR or DISR) to determine the appropriate current version of the SOAP standard to use. Examples in this document use SOAP 1.2.

2.1.2.2 Body

The `soap:Body` MUST be empty.

2.1.2.3 Example

An example of a SOAP message constituting a Retrieve Service request is shown in Figure 2. This example shows a retrieve with a URI of “urn:uuid:1225c695-cfb8-4ebb-aaaa-6fda344efa6a”. In addition, Figure 2 shows a notional example of Retrieve Properties.

```

<soap:Envelope>
  <soap:Header>
    <wsa:Action>urn:cdr:retrieve:2.0</wsa:Action>
    <cdr:Identifier>
      urn:uuid:1225c695-cfb8-4ebb-aaaa-6fda344efa6a
    </cdr:Identifier>
    <cdr:RetrieveProperties>
      <abc:RetrievalFormat>pdf</abc:RetrievalFormat>
    </cdr:RetrieveProperties>
  </soap:Header>
  <soap:Body/>
</soap:Envelope>

```

Figure 2: Example Retrieve Input

2.1.3 Output

The output of the CDR Retrieve Service is the content resource specified by the identifier provided in the service request. For requests that result in an error, a SOAP fault message will be output.

2.1.3.1 Header

The header of the SOAP message must contain the action element, as defined in WS-Addressing. The purpose of this element is to convey to the receiver which behavior was invoked.

Table 3: Required Header Elements for Retrieve Responses

Element Name	Description	Support
<code>/wsa:Action</code>	This element (content is of type <code>xs:anyURI</code>) conveys the value of the [action] property and indicates to a web service which operation should be invoked.	MUST be provided by the Service with a value of <code>urn:cdr:receive:2.0</code>

Additional elements, such as other WS-Addressing elements, MAY be added to the SOAP header.

2.1.3.2 Body

The SOAP body MUST consist of the retrieved content resource. Binary data MUST be base64³ encoded before inserting into the SOAP body.

³ Base64 is an encoding scheme to represent binary data in a ASCII format [BASE64].

To further reduce client and server processing when encoding or decoding large amounts of data, the SOAP Message Transmission Optimization Mechanism (MTOM) can and SHOULD be leveraged.

2.1.3.3 Output Example

This example represents sample output embedding the content resource (binary data) into the body of a SOAP message.

```
<soap:Envelope>
  <soap:Header>
    <wsa:Action>urn:cdr:receive:2.0</wsa:Action>
  </soap:Header>
  <soap:Body>
    IkdpmUgbWUgYSBsZXZlciBsb25nIGVub3VnaCBhbmQgYSBmdWxjcnVtI
    G9uIHdoawNoIHRvIHBSYWNlIGl0LCBhbmQgSSBzaGFsbCBtb3ZlIHROZS
    B3b3JsZC4iIC0gIEFyY2hpbWVkZXM=
  </soap:Body>
</soap:Envelope>
```

Figure 3: Example Retrieve Output

Figure 4 represents the same sample output as Figure 3, but with MTOM enabled. In this example, the output is a multipart mime message where the SOAP message is contained in the first part and references the content resource (raw binary data) is in the second part.

```
...
Content-type: multipart/related;
start="<1111111111111111@example.gov>";
type="application/xop+xml";
boundary="MIMEBoundary000000";

--MIMEBoundary000000
Content-Id: <1111111111111111@example.gov>
Content-Type: application/xop+xml;charset=utf-8;type="text/xml"
Content-Transfer-Encoding: binary

<?xml version="1.0" ?>
<soap:Envelope>
  <soap:Header>
    <wsa:Action>urn:cdr:receive:2.0</wsa:Action>
  </soap:Header>
  <soap:Body>
    <xop:Include href="cid:22222222222222@example.gov"/>
  </soap:Body>
</soap:Envelope>
--MIMEBoundary000000
content-id: <22222222222222@example.gov>
content-type: application/octet-stream
content-transfer-encoding: binary

...Binary Data...

--MIMEBoundary000000--
```

Figure 4: Example Retrieve Output (MTOM)

2.1.4 Post-Conditions

The following conditions MUST be met upon successful completion of the function.

1. The results returned are the content resource identified by the request.

2. The use this function has been audited according to applicable policy.⁴

2.1.5 Fault Conditions

An implementation of the Retrieve function MAY provide any of the faults listed in Table 4 as a SOAP Fault to the consumer.

An example of the Retrieve SOAP fault is shown in Figure 5. The fault response adheres to the SOAP 1.2 specification [SOAP]. The `/soap:Fault/Code/Value` element is used to convey the general type of error condition and MUST be from the enumeration `/soap:faultCodeEnum` as described in the SOAP 1.2 specification section 5.4.6. In addition, for the current use of this function, the fault MUST also contain the `/soap:Fault/Code/Subcode` element and its child element `/soap:Fault/Code/Subcode/Value`. The `/soap:Fault/Code/Subcode/Value` as listed in Table 4 supports automated processing of CDR specific errors. The `/soap:Fault/Reason` element has one or more `/soap:Fault/Reason/Text` elements as its children, where the value as listed in Table 4 for each `/soap:Fault/Reason/Text` element should be used to provide a human-readable explanation of the fault. The `/soap:Fault/Reason/Text` element MUST include the `xml:lang` attribute.

The following table outlines the Retrieve specific fault conditions that MAY be generated by an implementation.

Table 4: List of Faults

<code>/soap:Fault/Code/Value</code> <code>/soap:Fault/Code/Subcode/Value</code> <code>/soap:Fault/Reason/Text</code>	Fault Description
<code>soap:Sender</code> <code>cdr:retrieve:soap:fault:security</code> Security	The Consumer is either not authenticated or not authorized to perform the retrieve.
<code>soap:Sender</code> <code>cdr:retrieve:soap:fault:execution</code> Identifier Execution Fault	The Retrieve Component encounters an error retrieving the specified resource.

⁴ The use of this function may be audited according to applicable policy and may include auditing of the success or failure of the function.

2.1.5.1 Fault Message Example

The following shows a fault message of type “Identifier Execution Fault”.

```
<soap:Envelope>
  <soap:Header>
    <wsa:Action>http://www.w3.org/2005/08/addressing/fault</wsa:Action>
  </soap:Header>
  <soap:Body>
    <soap:Fault>
      <soap:Code>
        <soap:Value>soap:Sender</soap:Value>
        <soap:Subcode>
          <soap:Value>cdr:retrieve:soap:fault:execution</soap:Value>
        </soap:Subcode>
      </soap:Code>
      <soap:Reason>
        <soap:Text xml:lang="en">Identifier Execution Fault</soap:Text>
      </soap:Reason>
    </soap:Fault>
  </soap:Body>
</soap:Envelope>
```

Figure 5: Example Retrieve SOAP Fault

References

[BASE64]

“The Base16, Base32, and Base64 Data Encoding.” 2006.

[CDR-RA]

“CDR IPT Reference Architecture”, 1.1, 25 Feb 2011.

[CDR-RR]

“IC/DoD Content Discovery & Retrieval REST Interface Specification for CDR Retrieve 2.0.” 2012.

[CDR-SF]

“IC/DoD Content Discovery & Retrieval Specification Framework 2.0.” 2011.

[SOAP]

“SOAP Version 1.2 Part 1: Messaging Framework.” 2003. Available at
<http://www.w3.org/TR/2003/REC-soap12-part1-20030624/>.

[WS-A]

“Web Services Addressing 1.0 - Core, M. Gudgin.” 2006. Available at
<http://www.w3.org/TR/2006/REC-ws-addr-core-20060509>.

Appendix A. Mapping to Specification Framework

This section explicitly ties the items in this specification to the requirements of the CDR-SF. The CDR-SF identifies the requirements for creating specifications, while the implementation details are outlined here.

A.1. Retrieve Request

Table 5: Retrieve Request Mapping to Specification Framework

Specification Framework Variable	SOAP Retrieve Specification
Identifier	/cdr:Identifier
Retrieve Properties	/cdr:RetrieveProperties

A.2. Retrieve Response

Table 6: Retrieve Response Mapping to Specification Framework

Specification Framework Variable	SOAP Retrieve Specification
Content Resource	/soap:Body/{any}

SEQ Table